

## Rationality as the Rule of Reason

Antti Kauppinen ([antti.kauppinen@helsinki.fi](mailto:antti.kauppinen@helsinki.fi))

Forthcoming in *Noûs*.

In this paper, I propose that rationality consists in the rule of reason, where reason is understood in a quasi-Aristotelian sense as the set of capacities for recognizing favoring relations. To use John Broome's apt phrase, rationality is a matter of having one's mind in proper order. My claim is, roughly, that for our mind to be in proper order, we must form a competent take on reasons on the basis of available evidence and have the attitudes that correspond to it.

I will focus here on practical rationality, though I believe the same model can be extended to theoretical rationality. The problem I'll address can be seen clearly by considering the following three familiar cases of practical irrationality together:

### **Anna**

Anna must choose between having an operation today or tomorrow. She knows that an operation today would be mildly painful for five minutes, while an operation tomorrow would be mildly painful for seven minutes. Anna believes that she has most reason to choose to have operation today. Alas, she forms the intention to have the operation tomorrow.

### **Beate**

Beate intends to buy new shoes today, and believes that to do so, she must leave right now. Alas, she doesn't intend to leave right now.

### **Cecile**

Cecile must choose between having an operation today or tomorrow. She knows that an operation today would be mildly painful for five minutes, while an operation tomorrow would involve an hour of agony. Alas, tomorrow is Tuesday, and Cecile believes that she has no reason to avoid pain on future Tuesdays. (She also knows that when tomorrow comes, she will wish she had chosen the less painful operation today.) Since she believes she does have reason to avoid pain today, she believes she ought to choose tomorrow's operation, and intends to do so. Neither this intention nor the underlying beliefs clash with her other attitudes.<sup>1</sup>

Anna, Beate, and Cecile are all rationally criticizable for their intentions. Some might think that Cecile isn't irrational, but I agree with Derek Parfit (2011) that insofar as we use the term in its ordinary sense, we can't say that Anna is irrational but Cecile isn't – after all, her choice is much more foolish.

The best kind of theory of rationality would give a unified explanation of all these cases. But it is not trivial to do so. The scenarios pull in different directions: Anna's case highlights the rational significance of our own take on reasons, Beate's suggests that it is a kind of formal coherence that matters for rationality, and Cecile's indicates that rationality requires some sensitivity to genuine reasons, at least when the pertinent facts are known. Different theorists of rationality take different intuitions as their starting point. Some hold that irrationality is a matter of going against one's beliefs about (right kind of) reasons for attitudes. That's what Anna does, but it seems we can judge Beate to be irrational whatever she believes about reasons, while Cecile's irrational intention actually matches her beliefs about reasons. Others hold that to be

---

<sup>1</sup> The cases of Anna and Cecile are based on Derek Parfit's scenarios (2011, 120).

rational, we must conform to formal (coherence) requirements of rationality, which Anna and Beate violate. But they say nothing about Cecile, whose thoughts conform to such principles. Both above kinds of view also struggle to explain the point of being rational, since neither following our own take on reasons nor being coherent has a necessary connection to doing what we actually *should* do. This last problem is avoided by views according to which rationality is a matter of responding to reasons or evidence we possess. The challenge for such views, in turn, is to explain how rationality's demands can come apart from what our reasons or evidence support, and explain why at least some sorts incoherence indicate irrationality regardless of the correctness of our response.

What I propose here is a new way of reconciling these different demands. I maintain, roughly, that what rationality requires of us is forming competent perceptions of reasons on the basis of our evidence, and adopting the corresponding attitudes – for short, being ruled by our Reason. Very roughly, competent perceptions of reasons are implicit reason-takings that are fallibly sensitive to genuine reasons and patterns of reasons, relative to our evidence regarding non-normative facts. This view straightforwardly accounts for Anna's case, since she forms an intention against her competent perception of reasons. Beate's case, too, can be captured, but only if intentions for ends necessarily involve perceptions of reasons to perform actions one believes to be necessary to reach the end, and if such perceptions are competent. I'll argue that this is the case, so that Beate's failure to form the intention to leave is also irrational on my account. Finally, on my view, Cecile's rational failure is forming an attitude on the basis of an incompetent perception of reasons, in spite of knowing the pertinent facts. So if rationality consists in the rule of Reason, we have a desirable unified explanation of diverse cases of both structural and non-structural rationality and irrationality. This view also promises to explain the

distinctive authority of rationality, or so I'll argue: being ruled by our Reason is an authoritative ideal that grounds a kind of aretaic criticism for failures.

### **1. Rationality and Perceived Reasons (Or: What's Wrong With Anna, Tentatively)**

The cases I discussed in the introduction suggest that if our attitudes fail to fit together or match what our evidence supports, there is *prima facie* something wrong with us from the perspective of rationality. (It might be that this is not always the case – further conditions might need to be met, as I'll argue below.) Explaining why this is the case is one important desideratum for a theory of rationality. Another important desideratum is making sense of the fact that if we violate the demands of rationality, we're in some way *criticizable*. Unlike the demands of etiquette or Catholicism, rationality's demands seem to be somehow *authoritative* for anyone capable of meeting them, regardless of their desires or commitments. If I say that your intention to buy a fast car even though you're a very slow and accident-prone driver is irrational, I'm not just observing that there is a system of norms that happens to prohibit such an attitude. I'm saying that your intention manifests some sort of defect that you should care about. Many hold that rationality is *deontically* significant – that if rationality tells you to respond with F, you ought to respond with F (weak normativity) or that if rationality tells you to respond with F, you ought to respond with F *because* rationality tells you to do so (strong normativity).<sup>2</sup> I will be denying deontic significance in the final section. Nevertheless, among other things, a theory of rationality should explain why it is a criticism to call someone irrational, in particular a criticism that can be *addressed* to a person (rather than the sort of criticism one might make of a malfunctioning piece

---

<sup>2</sup> I use Broome's (2013) terminology here.

of equipment). Insofar as rationality really has some authority, a theory should also yield a good answer to the first-personal question “Why should I be rational?”.

These days, many think that rationality consists fundamentally in our attitudes fitting together, so that what it requires is a kind of coherence, for example the absence of contradictory beliefs. There has been much debate of late about the proper form of such structural requirements of rationality. Some think they have *narrow scope*, so that, for example, we’re rationally required to intend to F if we believe we ought to F. Others claim they have *wide scope*, so that rationality requires us to *either* not believe that we ought to F *or* intend to F. Both types of structuralist view could in principle explain what’s wrong with the cases of Anna and Beate. Yet intensive recent discussions have shown that each faces significant challenges. For example, narrow-scope views license bootstrapping rational requirements into being just by adopting any random attitude, possibly creating conflicting or mutually undermining demands (Broome 2013, Brunero 2010), while wide-scope views seem to permit wishful thinking (Schroeder 2009) and struggle to provide guidance to form attitudes (Kiesewetter 2017). Further, both kinds of view have problems with explaining the seeming authority of rationality – why be coherent when that may easily result in doing things you shouldn’t do, or in having beliefs that are not supported by your evidence? And of course, neither view explains what’s wrong with Cecile, whose attitudes are (by stipulation) perfectly coherent. I will not attempt to resolve these complex debates regarding putative requirements of structural rationality in this paper, but will rather focus on constructing a plausible alternative approach that postulates no underived structural or coherence requirements of rationality.

The kind of non-structuralist view of rationality that I will develop begins with the simple thought that we are the kind of creatures who take a stance on what we have reason to think and

do. In doing so, we ourselves set standards for which particular attitudes are required for a ‘proper order of our mind’. We’re then capable of responding to the reasons we perceive (in the non-factive sense) by forming the corresponding attitudes, though we sometimes fail to do so. As a rough approximation, rationality calls us to live up to our stance on reasons. (As Section 3 shows, this is only the beginning, not the whole story.) This alternative approach in its most straightforward form is most famously defended by Thomas Scanlon (1998, 2007) and Niko Kolodny (2005, 2007). For example, Scanlon says that “Irrationality in the clearest sense occurs when a person’s attitudes fail to conform to his or her own judgments: when, for example ... a person fails to form and act on an intention to do something even though he or she judges there to be overwhelmingly good reason to do it.” (1998, 25)

Setting aside for the purposes of this paper important issues regarding the nature and significance of the distinction between synchronic and diachronic demands of rationality, I’ll focus on the following simple formulation, which should suffice for the issues I examine:

*Rationality as Responsiveness to Perceived Reasons (RRPR)*

Rationality requires *S* to have attitude *A* if and only if *S* takes herself to have conclusive reason to have *A*.

Note that unlike Scanlon and Kolodny, I formulate the principle in terms of *taking oneself to have reason* rather than *belief* or *judgment* about reasons. This is because I don’t believe we have enough beliefs about reasons for attitudes to account for all instances of rationality, and even creatures that lack the concept of a reason can plausibly be irrational. But someone who lacks explicit beliefs might still take or perceive something as a reason. What is it to take oneself to have reason, then? On the view of normative reasons I have defended elsewhere (Kauppinen

2015), we respond correctly to a reason R for F-ing by thinking in an F-friendly way when doing something F-relevant in the presence of R. This includes the various ways we can be causally influenced in the direction of F-ing by thoughts of R – being motivated to F, assigning F-ing a higher utility in deliberation in the presence of R, raising the deliberative salience of F, and acquiring a disposition to do any of the previous. All these are consistent with not F-ing, if one has more reason to do something incompatible. It's a natural extension of this view to say that *taking* oneself to have reason is *endorsing* such influence, which might consist in something as minimal as willingness to appeal to R to defend thinking in a F-friendly way, if challenged by others, or a disposition to keep thinking in a F-friendly way if it becomes salient that I do so because of R, or a disposition to be puzzled or critical of others for failing to think in a F-friendly way when they're aware of R. This doesn't require having the *concept* of a reason – unless, of course, we take such dispositions themselves to be constitutive of concept-possession.

However we understand reason-takings, it might legitimately be asked why they should be thought to be so important for rationality. Why not say that it suffices for rationality that our attitudes are coherent or responsive to objective reasons? (After all, Beate and Cecile's irrationality seems to stem from some failure along these lines rather than going against their verdict on reasons.) Perhaps the most important reason to focus on perceptions of reasons is that rationality is a *person-level concept*. When we're criticized for irrationality, it is we *ourselves* who are taken to task for our responses. There's something amiss in the way in which we govern ourselves. Conversely, we might maintain coherence or reason-responsiveness by chance or by way of subpersonal processes, but that would not be to our credit as rational thinkers and agents (though it may be to our credit otherwise). As Ernest Sosa puts it in a different context, it is not to the credit of a ballerina if her graceful movements on the stage are caused by stumbling around

under the influence of a drug rather than her skill and artistic excellence (2007, 75–76). Similarly, if you acquire an intention for a necessary means to your end by random misfiring of neurons in your brain, it doesn't suffice to make *you* rational.

To put it differently, what is distinctive of us as specifically *rational* animals is that we're capable of *responding* to reasons *as* reasons (at least in the implicit fashion just described), not only *reacting* to reasons. In doing so, we take a normative stance on our own actions and attitudes, and become apt targets of rational criticism – the kind of criticism that doesn't have a point when it comes to creatures who are incapable of taking even an implicit a stance on reasons and responding to it. I emphasize that such criticism is in the first instance targeted at the subject. Here it may be useful to consider a parallel with epistemic justification. It is common to distinguish between questions of doxastic justification (whether a belief is justified) and personal justification (whether a subject is justified in believing it) (e.g. Littlejohn 2012). Rationality, on my view, is akin to (if not an element of) personal justification: calling a belief or intention of mine irrational is only shorthand for *my* being irrational in holding it. On this view, then, incoherence and failure to respond to evidence, for example, are not in themselves constitutive of irrationality, but only indications that somewhere down the line, our normative self-governance has failed.

So there's at least some good reason to think that rationality is linked to reason-takings of some sort. However, as John Broome (2007, 2013) persuasively argues, there are significant challenges to a view like RRPR, even if we set aside cases like Cecile's. First, by itself, it doesn't say anything about structural rationality (call this the Problem of Deriving Structural Rationality). Indeed, as Broome notes, even though RRPR is a self-governance requirement, it doesn't by itself entail even rational requirements that seem to be closely related. These include the roughly



correct principles that if you are rational, then, if you believe your evidence shows that  $p$ , you believe  $p$  (which Broome labels the Evidential Condition), and that if you are rational, then, if you believe your reasons require you to  $G$ , you intend to  $G$  (the Enkratic Condition) (Broome 2013, 98). And that's only the beginning, since any plausible account of rationality must also explain the irrationality of having incoherent attitudes, such as Beate's failure to intend the means to her end, which do not seem to involve thoughts about reasons at all.

Broome's second challenge is what I'll call the Problem of Independent Standards. It seems that rationality can require us to do something even if we *don't* take ourselves to have reason for it – maybe we're irrational in part precisely because we don't take ourselves to have reason to do certain things. Here's one of Broome's examples: "There is obviously something irrational about a person who does not intend a means she believes is necessary to an end she intends, even if she herself does not believe her reasons require her to intend a means she believes is necessary to an end she intends." (Broome 2013, 93) This is what happens in Beate's case – as I prefer to put it, we don't need to know about Beate's beliefs about reasons to know that she is (at least locally) irrational. Finally, specifically against Scanlon's account, Broome points out that people rarely have the kind of beliefs about reasons Scanlon thinks are crucial for rationality. He says that "The problem is that the antecedent in Scanlon's condition, that you believe your object-given reasons require you to  $F$ , is rarely satisfied. People rarely have beliefs of this sort" (ibid., 97). I'll call this the Problem of Rare Application.

## **2. Normative Constitutivism and Structural Rationality (Or: What's Wrong with Beate, Tentatively)**

The common assumption behind Broome's challenges is that all reason-takings are *optional*. The alternative I will explore is that taking oneself to have reason for further attitudes is partially *constitutive* of the relevant attitudes. Call this alternative view Normative Constitutivism. If it's true, it may be that all agents merely as believers and intenders perceive themselves to have reasons that rule out incoherent combinations. It turns out that Beate's irrationality is of the very same kind as Anna's (bearing in mind, again, that going against one's perceptions of reasons is only the beginning of the story to be completed in Section 3).

In this paper, I'll restrict myself to two attitudes that are crucial for explaining *practical* structural rationality: intention and belief about reasons for action. First, though, I will prepare the ground by saying a few words about belief in evidence. It might seem that the belief that E is evidence for *p* involves no thought about reasons. But to believe that E is evidence for *p* is obviously not merely to believe that E obtains. Nor is it to believe that E makes *p* more likely. That's just a belief about probability, not about evidence. You might think that the fact that you just spilled chocolate milk on your car seat makes it more likely that the car will smell bad tomorrow without yet thinking that it is evidence for the car smelling bad tomorrow. This is unsurprising, given that it is commonly thought that evidence is a normative rather than non-normative notion – as Jaegwon Kim puts it, “the concept of evidence is inseparable from that of justification” (Kim 1988, 390). Thomas Kelly (2014) goes so far as to suggest that insofar as evidence is what confers justification, the terms ‘reason to believe’ and ‘evidence’ are roughly synonymous. So it is plausible that for you to believe that E is evidence for *p* is, at least, for you to believe that (or at least take that) *E* is a reason to believe that *p*, and to not believe that not-*p*. (Those who believe in wrong kind of reasons for attitudes, should add ‘*object-given* reason’ or ‘truth-related reason’ here.)

So roughly, what is distinctive about believing something is *evidence* for something else is precisely taking it to provide *reason* for believing the other thing. I think something quite similar is true of belief in reasons for action. Just like to believe that E is evidence for truth of the proposition that *p* is, in part, to believe (or at least take) that E is a reason for an attitude, to believe that R is a reason to *make true* the proposition that *p* is, in part, to believe (or at least take) that R is a reason for an attitude. In the former case, the psychological state is belief, in the latter case, it is *intention*, since that's the attitude of being committed to make true the proposition. Why think so? To begin with, consider the oddity or even paradoxicality of saying "I have reason to F, but I don't have reason to intend to F". This makes sense only in the special context in which it is believed that you will F whether or not you intend to F, or that you won't F whether you intend to or not – for short, when acting isn't up to you, as Broome (2013) says. This oddity is explained if the following is true:

Action-Intention Principle: If you believe that *R* is conclusive reason for you to *F* and you take F-ing to be up to you, you take (or believe) *R* to be a conclusive reason for you to intend to *F*.

Note that the thesis is not that reasons for acting and reasons for intending necessarily coincide. In the Toxin Puzzle (Kavka 1983), for example, you may have a reason for intending that is not a reason for action. (I'm skeptical of the existence of such 'wrong kind of reasons' for attitudes, as opposed to right kind of reasons for wanting to have an attitude, but I'll leave this issue aside here.) My claim is just that when you take yourself to have reason to act and that your acting is up to you, you take yourself to have reason to intend, too. Here it may be tempting to add "insofar as you're rational". But this is not an option, if the project is to make use of the Action-

Intention Principle in understanding rationality.<sup>3</sup> The claim is a constitutive one: whether you're rational or not, you just don't think that you have conclusive reason to go to Paris while believing that it's up to you whether you go, if you don't think you have conclusive reason to intend to go to Paris. Here we might employ a test inspired by Moore's Open Question Argument: someone who says "I have conclusive reason to go to Paris and I won't go to Paris unless I intend to do so, but do I have conclusive reason to intend to go to Paris?" manifests at best a partial grasp of the concept of a reason for action. Perhaps she is confusing a reason to *go* to Paris (a reason to act) with a reason to *be* in Paris, which isn't similarly linked to a reason for intending.

So far, I've talked about how beliefs about reasons of one sort involve perceptions of other kinds of reasons. But how do we get from perceived reasons to rational requirements governing first-order attitudes that do not seem to involve beliefs about reasons at all? Since I focus on practical rationality, the nature of intention will be crucial to my story. What distinguishes intention from other pro-attitudes, such as desire? One difference is that intentions concern our own actions, while desires range more broadly. But there are also first-personal, action-directed desires. The key difference between such desires and intentions, I believe, is that only the latter involve *commitment* to act (see e.g. Bratman 1987). When we intend to act, we've made up our mind. My claim is that this commitment involves at least one kind of normative thought: that we have reason to do things that facilitate the realization of our aim, and to avoid doing things that stand in the way of its realization, including realizing incompatible goals.

---

<sup>3</sup> This is the most significant difference between the present (tentative) view and the account of structural rationality in Scanlon (2007). While Scanlon initially seems to consider something like normative constitutivism, he falls back on a view according to which it is "*irrational* for someone who has decided to do *A* at *t* (and has not changed his or her mind about this) to refuse to treat the fact that *B* would facilitate this as a reason for doing *B*" (Scanlon 2007, 93, my emphasis). But this amounts to giving up on his stated strategy of explaining structural irrationality in terms of what one considers oneself to have reason for. Perceptions of reasons couldn't be rationally required by further perceptions of reasons for them (a regress looms this way), so for them to be rationally required, Scanlon must postulate an independent and unexplained principle of rationality.

An example may help clarify this no doubt contentious claim. Suppose you know in the back of your mind that you can't get ice cream without opening the freezer door. You're contemplating whether to stick to your diet or get some ice cream, and settle on the ice cream option. Now the previously inert (perceived) fact that you must open the freezer to get ice cream appears to you as a reason to open the freezer, indeed as a conclusive reason to do so. It's not usual, to be sure, that you explicitly have the thought "What I have most reason to do right now is open the freezer". So why should we attribute the perception of a conclusive reason to you? To begin with, when you do open the freezer door, you're performing an intentional action, an action for reasons. Your reason for opening the door is that you must do so to get the ice cream. This fact doesn't simply motivate you (it's not a mere "motivational reason" in that sense), but it's something that you take to favor or justify your action. You endorse yourself being moved to act in this way by the fact. As I said above, this is manifest, among other things, in your dispositions to defend thinking in a freezer-opening-friendly way, given that it is necessary to get ice cream, to keep thinking that way when your attention is drawn to the reason, or criticize others in a similar situation who fail to think that way. So if someone says "You shouldn't open the freezer!", then, as long as you don't give up your intention to get the ice cream, the natural thing for you to say in your defense is "But I need to open it to get ice cream!". Or if someone asks you "Why are you opening the freezer door?" as you're doing it, you might pause to reflect and say "Otherwise I won't get ice cream", and continue opening the door, in contrast to cases in which you stop what you're doing, because you don't endorse the reason for which you act. And if I say "I'm going to get some ice cream now", but make no effort to open the freezer, you're at the very least puzzled, and might charitably conclude that I don't really intend to get ice cream.

The underlying explanation for this is that intentions are (at least potentially) a means of self-control and self-governance, and because we're agents who respond to reasons we perceive, intentions work (at least in part) by way of affecting our perceptions of reasons. This may be obscured in situations in which other reasons are more salient. Suppose I judge that I really ought not eat ice cream tonight, because I have a photo shoot for a bikini calendar tomorrow and don't want to appear bloated. But let's assume that temptation wins, and I nevertheless form the *akratic intention* to eat ice cream. Now, I might realize that not only will I not eat ice cream unless I open the freezer, but also that I will if I do. Consequently, I may think that I shouldn't open the freezer, since I shouldn't eat ice cream. So it might seem that I don't (or need not) think that I have conclusive reason to open the freezer. But I maintain that I do, insofar as I'm genuinely settled on eating ice cream.

Here a separation of different kinds of 'ought' or requirement may clarify the situation. I may well think that I *prudentially* ought not open the freezer door. Still, if I have, against my perception of the balance of reasons in the situation, settled the deliberative question in favor of getting ice cream by forming the intention to do so, opening the door will appear to me as *the thing to do*, given its apparent necessity for doing what I'm going to do. If I do it, I'll do it for that reason. My contrary perception of prudential reasons may mask at least some of the dispositions to defend and criticize (though not the disposition to continue when my attention is drawn to my reason for acting). I might say "Man, I really shouldn't open this door now," meaning I think doing so is not in my best interest. However, as a reflective rational agent, I nevertheless exert control over means-actions by way of my take on reasons for them. It's not an open deliberative question for me whether to open the freezer or not, as long as I intend to eat ice cream, so my talk of what I have or don't have reason for is mere talk in the sense that it doesn't

exhaust my take on reasons. When I've settled what I'm *going to* do, I've also settled what *to* do in important downstream respects.

Here's one way of spelling out this aspect of intention:

Intention Commitment (rough): In intending to *E*, we take the perceived fact that *M*-ing non-trivially facilitates *E*-ing to be a reason (of a strength proportionate to the degree of facilitation) for us to *M*, and the perceived incompatibility of *E*-ing and *F*-ing as a conclusive reason not to *F*.

I say this formulation is only rough, since it leaves unspecified the relative times of the end- and means-actions, which are crucial for explaining diachronic rationality. Again, since the focus of this paper is on the more fundamental issues regarding the relationship between reasons and rationality, I must leave discussion of diachronic rationality for another occasion.

### *Structural Rationality*

I'm next going to argue that if (and only if) something like the Normative Constitutivist theses I've just defended are true, a perceived reasons account like RRPR can after all meet the challenges that Broome raises. I'll start with the Problem of Deriving Structural Rationality. Why does incoherence indicate (and in some cases even guarantee) that we are responding with attitudes we take ourselves to have reason not to have? To illustrate the solution, I'll first show how the irrationality of akrasia can be accounted for in terms of perceptions of reasons. Here's how the assumption that someone is akratic results in her not being fully rational on this picture:

1. S believes she has conclusive reason to F, and that F-ing is up to her, but doesn't intend to F. (Akrasia)

2. S takes herself to have conclusive reason to intend to F. (from 1, Action-Intention Principle)
3. Rationality requires S to have attitude A if S takes herself to have conclusive reason to have A. (RRPR)
4. So, rationality requires S to intend to F. (2, 3)
5. So, S is (at least) not fully rational, because rationality requires her to intend to F, and she doesn't intend to F. (from 1, 4)

This derivation should be quite self-explanatory. Since the Action-Intention Principle says beliefs about reasons for action entail perceptions about reasons for intentions, it is evident why akrasia turns out to be problematic. Note, in particular, that with these assumptions, there is no need for an independent (wide or narrow scope) principle of structural rationality that says akrasia is irrational. It's just that whenever you believe yourself to have conclusive reason to do something that's up to you, you also take yourself to have conclusive reason to intend to do it, and hence go against your own take on reasons for attitudes if you fail to form the intention.

Other platitudes about structural rationality would take a bit more work to explain. It will hopefully suffice to illustrate how this is done to discuss the following platitude:

*Instrumental Failure:* If you intend to E, believe that you will not E unless you M and you will E if you M, believe that your M-ing is up to you, and don't intend to M, you're not fully rational.

If we focus on the simplest case of necessary and sufficient means, the derivation goes as follows:



1. S intends to E, believes that she will not E unless she M-s and that she will E if she M-s, and that M-ing is up to her, and does not intend to M. (Instrumental Failure)
2. If S intends to E and believes that she will not E unless she M-s and that she will E if she M-s, S takes the necessity and sufficiency of M-ing for E-ing to be a conclusive reason for her to M. (By Intention Commitment)
3. So, S takes herself to have conclusive reason to M. (From 1 and 2)
4. If S takes herself to have conclusive reason to M while believing that M-ing is up to her, S takes herself to have conclusive reason to intend to M. (By the Action-Intention Principle).
5. So, S takes herself to have conclusive reason to intend to M. (From 3 and 4)
6. Rationality requires S to have an attitude if and only if she takes herself to have conclusive reason to have it. (RRPR)
7. So, rationality requires S to intend to M. (From 5 and 6)
8. So, S is not fully rational, because rationality requires S to intend to M, and S doesn't intend to M. (From 1 and 7)

If Normative Constitutivism is true, the steps are again very straightforward. Intention Commitment does a lot of work here, and when the means are believed to be necessary and sufficient, simple steps lead to a stringent rational requirement to intend the means. I don't claim that explaining instrumental rationality in general is simple at all – a lot of work clearly needs to be done to account for the rationality of taking the best non-necessary means, for example (see Kolodny forthcoming). Moreover, it may be rational to escape the requirement to take the means by giving up the end (see e.g. Lord 2011). Any synchronic account will only diagnose

irrationality without giving guidance for how to respond to it. However, since my purpose here is merely to illustrate how seeming coherence requirements can be derived from perceptions of reasons for attitudes, I will leave such issues for another occasion. What I've said here should suffice to show why Beate's irrationality is like Anna's: in spite of (possibly) lacking explicit beliefs about (right kind of) reasons for attitudes, Beate, too, fails to have an attitude (the means-intention) that she takes herself to have conclusive reason for (in virtue of having the end-intention), and is consequently irrational unless she either adopts the means-intention or drops the end-intention.

As I've said, in this paper I will focus on practical rationality. But it may be worth noting that there are various ways of extending the strategy to doxastic rationality. Belief, after all, is a commitment that is in some ways parallel to intention. It might be, for example, that part of what it is to believe that  $p$  is to take oneself to have sufficient reason to believe whatever follows from  $p$ , and conclusive reason not to believe what is incompatible with  $p$ . If that is the case, anyone who believes both  $p$  and not- $p$  will believe something she takes herself to have conclusive reason not to believe. And anyone who believes that  $p$  and that  $p$  entails  $q$  will take herself to have sufficient reason to believe that  $q$ , in which case RRPR requires her not to believe that not- $q$  and permits her to believe  $q$ , as long as she believes that  $p$  and that  $p$  entails  $q$ . (Again, as a synchronic view, it is neutral on which way to resolve the tension going forward.) But I will leave it open here what is the best way to cash out the commitment involved in belief in a way that accounts for structural rationality without resulting in paradoxes.

What these derivations suggest, then, is that when conjoined with Normative Constitutivism, a perceived reasons account of rationality can explain the structural irrationality of familiar cases like akrasia and failing to intend the means to one's end. Indeed, unless such

perceptions of reasons are part and parcel of intentions and beliefs, there's no way of systematically explaining why certain forms of incoherence indicate that one is irrational. That's why Normative Constitutivism is indispensable for any perceived reasons account of rationality.

So far, I've focused on the Problem of Deriving Structural Rationality. But it should also be clear now what the response to the Problem of Independent Standards is: regardless of their particular views on reasons, everyone will be subject to the certain structural demands of rationality, namely those that derive from perceptions of reasons that are constitutive of the relevant attitudes. Any agent who has intentions in the first place will perceive reasons to take the means to her ends, even if they hold crazy explicit beliefs about their reasons, so the kind of agent Broome envisages will be irrational by RRPR. Similarly, the Problem of Rare Application isn't an issue, since implicit perceptions of (right kind of) reasons for attitudes turn out to be relatively ubiquitous.

### **3. The Rule of Reason (Or: What's Wrong With Cecile – and Anna and Beate)**

With Normative Constitutivism, a perceived reasons view can robustly explain structural demands of rationality. But it can't explain why Cecile is irrational. Indeed, RRPR yields the verdict that Cecile is perfectly rational, even if she is substantively mistaken about her reasons. If Cecile's choice is irrational – and again, as Parfit says, she seems far more irrational than Anna – this is a problem for RRPR. Her mistake isn't just a theoretical one. I agree with Parfit that “if we have irrational beliefs about practical reasons, and about what we ought rationally to want or to do, our having such beliefs makes us in one way practically irrational” (Parfit 2011, 119).

But problems with perceived reasons accounts go deeper. Most importantly, it offers a highly deflationary “transparency account” of the *normativity* of rationality (Kolodny 2005). In

Scanlon's words, "Normativity enters only from the point of view of the person who has these attitudes, and therefore sees the relevant considerations as reasons." (2007, 87) To criticize someone for being irrational, on this account, is, in Kolodny's words, simply to make a "descriptive, psychological claim" that "from his point of view ... he has conclusive reason to have the attitude", that is, to call his attention to what he himself already thinks (Kolodny 2005, 455). But criticizing someone isn't just pointing out that she has violated her own standard, but assumes that the standard is *authoritative* for the target of criticism (Kiesewetter 2017). Sometimes the criticism is precisely that the target's standards are mistaken. So failure to meet one's own standards doesn't seem to capture the force of the charge of irrationality.

Both of these problems could be avoided if having a proper order in one's mind required responding correctly to the reasons there objectively are for us. Such reasons do not require us to do crazy things, and are genuinely normative for us, if anything is. Yet collapsing the distinction between requirements of rationality and requirements of reason comes at too great a cost. Failures of rationality are a *distinct* sort of normative mistake. For objectivists about reasons, it is possible for us to have a reason to do something even if we're not aware of it (or fail to recognize it as a reason). In the simplest case, if I falsely believe that there is a bear in my room, it may be rational for me to avoid going there, even if I actually have most reason to enter there (maybe it would be the only chance to meet the love of my life). In such cases, rationality and reasons point in different directions. Insofar as rationality is linked to reasons, the connection must be less direct.

Parfit's account of rationality is an attempt to forge such an indirect connection. Consider that if my belief that there is a bear in the room were true, I would have a reason not to enter. In Parfit's terms, when this is the case, I have a (merely) 'apparent reason' not to enter. (Some apparent reasons are, of course, genuine reasons.) According to him, "we ought rationally to

respond to apparent reasons even if, because our beliefs are false, these reasons are not real” (Parfit 2011, 111). So the demands of rationality are derived from the demands of reasons, but they can come apart. In a case like Cecile’s, the two kinds of demand coincide: not only does she have decisive reason to choose the less painful operation, but it is also rationally required for her to do so, since the truth-makers of her non-normative beliefs are decisive reasons to choose the less painful operation. This requirement of rationality is independent of her beliefs about reasons, though Parfit grants that she would be more rational if her normative beliefs were true.

On this picture, the term ‘apparent’ is in a way misleading, since it need not be apparent *to me* that I have a reason to do something – I only need to have a *non-normative* belief about something that actually is a reason for me, or would be, were the belief true. And this is a problem. If I don’t recognize that something would weigh in favor of an attitude, I need not be rationally criticizable for not forming the attitude. As Broome (2007) observes, I might know a non-normative fact about something, say that the chicken on my plate has salmonella, while non-culpably failing to recognize this a reason not to eat (maybe a trusted friend played a prank on me and convinced me salmonella is a delicious Italian fish sauce). He argues that in such a scenario, intending to eat is not irrational. His principled explanation for this is that rationality supervenes on the mind, not on how things actually are, normatively speaking. I’ll return to this issue below, but from my perspective, it suffices that rationality is a person-level concept. It’s not just a matter of tracking or failing to track what reasons we have, but rather supervenes in part on our first-personal take on reasons.

One way to take this into account is to say, as Errol Lord (forthcoming) does, that rationality consists in responding correctly to the reasons we *possess* (or have), where I possess a reason R to F if and only if R is a reason to F, I know that R obtains, *and* I treat R as a reason to F.

Since I don't in this sense possess the reason not to eat the chicken in Broome's case, Lord's view, unlike Parfit's, doesn't entail that it is irrational to eat it. But if objectivism about reasons is true, it still forges too tight a link between rationality and reasons, since I might be rationally required to act or have an attitude when I in fact lack sufficient reason, and *a fortiori* can't possess one. Suppose that Della knows she has promised Larry to return his book tonight, and treats that as a conclusive reason to return the book to tonight. It seems she's rationally at fault if she doesn't intend to return the book, even if there is as a matter of fact no conclusive reason for her to return it, and consequently no reason she could possess. After all, it's possible that in this case her having promised is not a conclusive reason to return the book – perhaps there's a significant cost to her involved and Larry doesn't want the book, and has forgotten all about the promise. It could also be that unbeknownst to Della, she *can't* return the book to Larry. Such inability plausibly functions as a disabler, so that the fact that she promised to return the book is not a reason to do so in this particular context. Assuming Della's reason to intend to return the book derives from her reason to return it, it, too, will be outweighed or absent. So even if she doesn't possess a conclusive reason to (intend to) return the book, she can be rationally required to (intend to) do so, given her (reasonably held) beliefs about reasons and knowledge of non-normative facts.

These arguments presuppose that there could be reasons for us to have attitudes or do things that are not available to us, or part of our evidence. Lord and Kieseewetter, among others, deny this, and argue for an informational constraint on ought and reasons, so that given her ignorance, Della ought to intend to return the book. I cannot enter into this debate here. But even if we grant that reasons are evidence-relative, the possibility of misleading higher-order evidence (Lasonen-Aarnio forthcoming, Worsnip forthcoming) shows there is a gap between reasons and

rationality. It appears to be possible that in the light of our evidence, we should respond with R, but at the same time, we have misleading evidence to the effect that our evidence does not sufficiently support responding with R.<sup>4</sup> Consequently, we faultlessly take it that we don't have sufficient reason to respond with R. It seems that it is not rational for us to adopt the attitude that our evidence in fact supports, when it goes against our well-supported but mistaken take on that evidence. Again, rationality's demands are linked to our take on reasons rather than what we have reason to do – *even if* the relevant reasons are restricted to those that are available to us.

### *Rationality and Competence with Reasons*

In short, it seems to me that in spite of recognizing the issue at stake, neither Parfit nor Lord and Kieseletter strike the right kind of balance between the *subjective* aspect of rationality as having the attitudes we take ourselves to have reason to have, and the *objective* aspect of rationality as responsiveness to objective reasons. I believe that the best way to navigate this thicket begins with the observation that our perceptions of reasons can be better or worse in the sense of resulting from a fallible *competence* with objective reasons, or failing to do so. This type of view has recently been sketched by Kurt Sylvan (2015). Although he frames his discussion somewhat differently, I believe it is well suited to solve the problems I have been discussing. Here is how he puts the main point:

[E]ven if we recognize a notion of rationality that falls between coherence and correctness, we do not need to think that rationality is divorced from objective correctness: it is a competence to achieve objective correctness, however fallible and unhelpful in

---

<sup>4</sup> I cannot here discuss possible objections and replies, but I find Worsnip's and Lasonen-Aarnio's account compelling, and will assume the possibility of mismatch between evidence and higher-order evidence.

unfortunate circumstances. ... [A] relevant reasons-sensitive competence is a competence to treat R-like considerations like objective reasons to do F-like things only if they are, when true, objective reasons to do F-like things. (Sylvan 2015, 605, 599)

In the rest of this paper, I will further develop this idea. To begin with, what does competence mean here? The relevant use of the term originates in (reliabilist) virtue epistemology. To put it crudely, while the reliabilist idea that knowledge is true belief produced by a reliable process has many attractions, it appears to generate false positives, classifying true beliefs as knowledge when they're not. Of particular interest here are cases that fail to amount to knowledge, because they're not genuinely achievements of the agent. The virtue epistemological idea, especially in the tradition started by Ernest Sosa, is to say (again very roughly) that true beliefs amount to knowledge when they manifest a reliable disposition, attributable to the agent herself, to get it right (in the relevant circumstances) – a competence of the agent. In Sosa's terminology, performances that aim at a goal in general are accurate or correct when they hit the relevant target, adroit or competent when they manifest a disposition, with a basis in the agent, to hit the target (in the relevant circumstances), and apt when correct because competent. Sosa and other virtue reliabilists like John Greco (2010) are particularly interested in aptness, since they hold that knowledge is apt belief, but my interest is rather in adroitness or competence, since unlike knowledge, rationality isn't a success notion.

What is competence with reasons? As a first pass, we might say that when we competently take something to be a reason, our reason-taking results from a disposition that tends to hit the target – a disposition to take R to be a reason to F only when R really is a reason to F if it obtains. My disposition to take the putative fact that the best-before date of the chicken in my



fridge is past to heavily count against eating it is a (part of) a reasons-sensitive competence, assuming there's reason for me not to eat chicken, if indeed it is past its best-before date. Of course, I might be mistaken in thinking I shouldn't eat the chicken even if I'm competent with reasons, if I'm mistaken about the facts – perhaps the appearance that the best-before date is past is an illusion generated by an evil demon. After all, whether I ought to F in the presence of R depends both on whether R-facts are a (conclusive) reason to F if they obtain, *and* on whether they actually obtain. A reason-sensitive competence only targets the first conjunct, the favoring relations between putative facts and actions or attitudes.<sup>5</sup>

It's not enough for competence with reasons in a domain, however, to recognize that something favors an attitude or action. Reasons, after all, are characteristically *pro tanto* – they favor something to some extent. They have a strength or *weight*. That it would be exhilarating for a while counts in favor of jumping off the roof, if only a little bit. That you would break your leg or worse counts heavily against it. To be competent with reasons, you must be disposed to recognize (at least roughly) such differences in weight and take them into account when forming a take on the balance of reasons. Insofar as other factors make a difference to the weight of reasons – if there are, for example, considerations that are not themselves reasons but attenuate or intensify their weight, or enable or disable them, as Jonathan Dancy (2004) argues – being competent with reasons requires being disposed to recognize them for what they are as well. Further yet, reasons often depend on one another, as Joseph Raz (2005) emphasizes. Your reason to go catch the train to Oxford depends on your reason to get to Oxford – if you lose one, you lose the other. Recognizing such *patterns* in reasons is thus another crucial part of our

---

<sup>5</sup> I return to the implications of this for rationality's supervenience on the mental in the next section.

competence with them (and plays a role in explaining what's wrong with incoherence, as we'll see).

I want to emphasize that the kind of reason-sensitive competence that is relevant to rationality is not just a matter of *reacting* to reasons. I take it that it's plausible that there are reasons for many creatures that are not apt to be assessed for rationality to respond to their changes in their environment in certain ways, and that they may react to those reasons appropriately enough. People, too, can react to reasons without responding to them *as* reasons, as might happen when a tennis player changes position in order to be able to catch a volley. When we do respond to something *as* a reason, we take it to be a reason in the way I've described above – we endorse its influence on us, are disposed to appeal to it in defending our attitudes or actions, and so on. Such recognition of reasons as reasons makes possible explicitly weighing them against each other, as well as the social practice of giving and asking for reasons. The former may help correct for biases in our thoughtless or intuitive responses, and the latter makes it possible to learn from others, and so improve and coordinate our responses. So it is no wonder it is what makes us subject to rational criticism and praise.

As Sylvan (2015) observes, there are many dispositions that amount to the relevant kind of competence. The inductive disposition to take an observed regularity to favor belief in a universal generalization is arguably one example on the theoretical side, as is the disposition to take visual appearances in daylight to constitute evidence for the color of things. In favorable conditions, these dispositions result in the correct take on reasons for belief. It is a question for substantive normative theory just which dispositions constitute competence in the practical domain. Presumably the tendency to take future agony for oneself or others to constitute a reason against an action is one such disposition. There is good reason to think that we can be competent

reason-recognizers in one domain without being such in all domains. This suggests that the broad notion of competence with reasons in reality fragments into many local competences.

Now, for rhetorical ease, I am going to use the label *Reason* with a capital R to refer to the set of our reason-sensitive competences. I grant that this is a broader conception of Reason than is common (though compare Raz 2011). Sometimes people use ‘reason’ exclusively as the name for an inferential capacity. But there is also precedent in the Aristotelian tradition for thinking of Reason as a capacity to recognize reasons in general. Given our different areas and levels of competence with reasons, we partake in Reason in different ways and to different degrees.

So, leaving out some qualifications that may well turn out to be warranted, the thesis I tentatively want to defend is the following:

*Rationality as the Rule of Reason (RRR)*

Rationality requires S to have attitude A if and only if S takes herself to have conclusive reason to have A, and does so competently.

Rationality does not permit S to have attitude A if and only if S takes a stance on her reasons regarding A, and either does not take herself to have sufficient reason to have A or does so, but incompetently.<sup>6</sup>

I will next spell out some implications of this view by examining the cases I’ve already introduced, and consider some possible qualifications. Let’s start with Cecile’s case. Her perception of reasons is incompetent, since she fails to recognize a strong reason against the very painful Tuesday operation, and consequently incompetently takes herself to have conclusive reason to take the Tuesday option. By the Action-Intention Principle spelled out in Section 2, she

---

<sup>6</sup> Alex Gregory deserves credit for helping with the present wording of the thesis.

also takes herself to have conclusive reason to intend to take the Tuesday option. Since her perception of reasons is incompetent, the first clause of RRR doesn't apply to her, and she's *not rationally required* to intend to take the Tuesday option (in contrast to RRPR and Scanlon's view). What is more, since Cecile takes a stance on her reasons regarding the operation, but only incompetently takes herself to have sufficient reason to have the operation on Tuesday, the second clause of RRR says that she is *irrational* in intending to have the operation on Tuesday.

I emphasize that the second clause of RRR, as formulated, says that *insofar as* we form a take on reasons for something, it must be a competent one, or we're not rationally permitted to have the attitude. The reason for this formulation is that in the absence of the qualification, we would be irrational whenever we have an attitude without competently considering reasons for it. But our perceptual beliefs, for example, seem to be typically held without taking a stance on reasons for them, and are nevertheless not irrational. (However, many of us have a more general stance on trusting our senses in certain situations, or treating their deliverances as evidence.<sup>7</sup>) It would perhaps be best to add a clause to RRR saying that attitudes that we genuinely have in the absence of any stance on reasons that pertains to them are *non-rational*. Insofar as they diverge from attitudes that *would* result from using Reason, we could say that they're *open to rational criticism*, even if not strictly irrational. Some might object here that this allows us to avoid irrationality too easily, simply by forming attitudes without taking a stance on reasons regarding them. At a first glance, simply forming an intention to have the painful Tuesday operation would be non-rational (and thus not irrational) on this view. But first, it's not clear that this is the wrong verdict (it would certainly be less obviously irrational than Cecile's intention), and second, for

---

<sup>7</sup> This is the tack taken by Scanlon, who holds that the formation of such unreflective attitudes is still constrained by our "general standing judgments about the adequacy of reasons" (1998, 24). While this may be true in many cases, it still seems too restrictive – surely children's perceptual beliefs aren't irrational even if they don't have such a general stance.

the vast majority of people, forming such a brute intention would go against their competent take on reasons to avoid future agony if they can, so that it would after all be irrational according to RRR. We could also go pragmatic and add a condition to the effect that if one has evidence bearing on whether to F (or whether  $p$  is the case), and one's evidence also suggests that F-ing or failing to F (or believing that  $p$ ) significantly impacts what one cares about, one is *rationally required to take a stance on reasons* for F-ing (or for believing that  $p$ ). Basically, if you have evidence that doing something might be a terrible idea from your perspective or that you could end up hurting yourself if you acted on an assumption, you can't rationally ignore it. This seems plausible, and consistent with the original motivations for RRR.

Another possibly warranted qualification to the second clause is introducing the notion of *minimal competence* with reasons, a set of dispositions to recognize at least the clearest and most obvious reasons that apply to one, to adequately weigh them against each other in arriving at a verdict about what I ought to do, and to rethink matters if we find ourselves perceiving sufficient reasons both for and against an attitude. We could then say that rationality prohibits attitudes that are not even minimally competently held (when one takes a stance on reasons). This would mean that while Cecile's intention is irrational, many other *mistakes* are not, as long as they are still minimally competent. For example, RRR modified along such lines doesn't entail that it is irrational to choose a holiday destination that is, in the light of one's evidence, all-things-considered somewhat less satisfying than a competing option— given that it is often difficult to correctly compare options that differ along many dimensions, the choice doesn't call into question one's at least minimal competence.<sup>8</sup>

---

<sup>8</sup> This means that a qualified RRR allows making the distinction between the inconsiderate, selfish, and the cruel, on the one hand and the irrational, on the other, as long as the reasons the

Cecile's case is obviously also captured by Parfit's and Lord's views. But note that unlike Parfit's view, RRR doesn't imply that it is irrational for me to intend to eat the salmonella chicken when I mistakenly think there's sufficient reason to eat it, as long as my mistaken perception of reasons is a competent one. That might well be the case, if it is based on misleading testimony of a trusted source. (If I incompetently believe salmonella is good for me, RRR implies it is indeed irrational for me to intend to eat the chicken.) And unlike Lord's view, RRR implies Della is rationally required to intend to return the book, even though she doesn't in fact possess a reason to do so, as long as she competently perceives her having promised as a conclusive reason. Since having made a promise to do something is indeed normally a strong reason, and (let's stipulate) Della is not aware of any stronger competing considerations or defeaters, Reason will in such a case tell her to return the book. Finally, when we have genuinely misleading higher-order evidence, we will competently take it that we ought to think or do something our available evidence doesn't in fact support, so that's what rationality calls us to do. So RRR yields the correct verdict about these cases, too.

#### **4. Challenges**

In this final section, I'll address three important challenges to the conception of rationality as the rule of reason.

##### *4.1 Supervenience on the Mental*

According to RRR, competence requires a kind of sensitivity to the reasons there actually are, not only internal consistency of some sort. This means I reject the popular thesis that rationality

---

morally vicious fail to recognize or respond to are non-obvious. Whether this would satisfy critics like Williams (1981) and Scanlon (1998), who insist on this distinction, is not clear.

supervenes on the mental. The standard argument for the thesis is a form of the New Evil Demon argument (Cohen 1984), originally directed against reliabilist theories of knowledge and justification. Here's a scenario:

Normal Norma and Envatted Enya have exactly the same experiences, apparent memories, and intuitions, as well as mental histories. The only difference is that while Norma's experiences tend to be veridical, Enya's are illusions generated by an evil demon.

Consequently, many of Enya's beliefs are false and intentions misguided.

It is widely agreed that Norma and Enya (or any two subjects alike in non-factive mental states and histories) are necessarily equally rational. Yet Enya's belief-forming processes are not reliable: in her case, believing that there is a red bus in front of her on the basis of a visual experience as of a red bus will result in a false belief. So, many conclude, rationality cannot depend on how things actually are – it must supervene on non-factive mental states (e.g. Wedgwood 2002).

This is a challenge for RRR. How could Enya's take on the world manifest a disposition to recognize reasons, if she systematically gets wrong what she has reason to believe and intend? If Norma and Enya are equally rational, rationality must depend on something that does not vary between them, and the obvious candidate is their non-factive mental states. But it is important to note that it is not *only* non-factive mental states that are constant between actual and demon-worlds. The same *necessary truths* will also hold. If Norma gets a mathematical truth right, then so does Enya, since their mental histories are identical, and  $2+2=4$  in both worlds. So if Norma is competent with mathematics, so is Enya, even though mathematical competence doesn't fully supervene on the mental. This is important, since *truths about basic favoring relations are*

*necessary truths*. This is widely accepted in the case of basic moral truths: if causing suffering basically counts against an act, it necessarily counts against it – there’s no possible world in which causing suffering doesn’t count against the act. There’s no reason to think other basic favoring relations, whatever they are, would be different. If it’s the case that F-ing’s being fun basically favors F-ing, then it does so in every possible world. So if Norma has a disposition to get basic favoring relations right, so does Enya. And if Norma’s taking herself to have (basic) reason to F manifests this competence, so does Enya’s corresponding take, and RRR says they’re equally rational.

Of course, not all favoring relations are basic. Norma takes the fact that Joe is having a party to favor going to Joe’s. But Joe’s having a party is a reason for her to go to Joe’s only if the party is fun, say. Her reason to go to the party derives from her reason to have fun, and the fact that the party is fun. Let’s stipulate that she recognizes that she has reason to do fun things and believes that the party would be fun in virtue of her experience with Joe’s earlier parties, and consequently takes herself to have reason to go. What about Enya? The demon could easily deceive her about Joe’s party. She could have the same (non-factive) experiences, same seeming memories, and reason the same way as Norma, and come to falsely believe she has a reason to go to Joe’s party. Yet it seems that Enya would be no less rational if she intended to go to the party on the basis of her take on reasons. This is a problem for RRR – how could Enya be as competent with reasons as Norma, if she systematically gets derivative favoring relations wrong?

It seems to me that the best answer to this challenge is to say that a reason-sensitive competence does not strictly speaking require reliably getting it right, but only reliably getting basic favoring relations right, *and* forming a take on derivative favoring relations on the basis of competence with basic favoring relations and empirical assumptions that are supported by the



totality of one's experience, or the way that things seem to one. For example, Enya recognizes that she has a reason to have fun, and thinks she has reason to go to the party on the basis of combining this take on reasons with the assumption, supported by the totality of her experience, that the party would be fun. Since her experience is systematically misleading, she is mistaken, but this doesn't call her competence with reasons into question – she reliably gets the basic favoring relations right, and forms her view of derivative reasons on the best available grounds.

There are several things worth noting about this view. First, while it entails that any subjects with identical non-factive mental states and histories are necessarily equally rational, it nevertheless still denies the supervenience of rationality on the mental. Rationality still requires competence with basic favoring relations, not just some sort of internal coherence. Suppose Norma and Enya are faced with a choice between two options that are otherwise equally appealing, except that A is much more fun than B, while B is a tiny bit more educational. If Norma and Enya fail to recognize that an option's being fun favors choosing it (to some extent), and take themselves to have conclusive reason to choose B, their choice of B may well fail to be rational, because it's based on an incompetent take on reasons. While they might be blameless for their choice, they're potentially still liable to rational criticism. Rationality isn't just blamelessness. (I'll return to this below.)

Second, the suggestion isn't, and doesn't purport to be, a solution to the New Evil Demon problem for externalist theories of justification. All I claim is that since some favoring relations are necessary truths, the new evil demon scenario doesn't rule out the possibility that rationality supervenes, in part, on the right sort of connection to extra-mental facts. Third, the view is neutral between internalist and externalist conceptions of doxastic justification and related views about evidence. Internalists would say that Norma and Enya are equally justified in their beliefs about

the world, and some of them would say that they have the same evidence. In these terms, the view would be that competence with derivative reasons requires competence with basic reasons and justified beliefs about pertinent contingent facts. Externalists, in turn, might deny that Norma and Enya have the same evidence (for example because evidence consists in what we know), or that Enya's empirical beliefs are doxastically justified.<sup>9</sup> But as Clayton Littlejohn (2012) argues, even such externalists can and should grant that Enya is *personally justified* in believing that the party would be fun, say, and that she's as rational as Norma. So in these terms, my thesis is that competence with reasons requires competence with basic favoring relations and combining opinions regarding basic reasons with personally justified beliefs about the world in forming a take on derivative reasons.

#### *4.2 Explaining Structural Rationality*

How does rationality as the rule of Reason account for irrationality of incoherent attitudes? What does it say about Anna's and Beate's cases? They are both instances of structural irrationality, and on the face of it, RRR doesn't seem to say anything about Beate, at least. Indeed, if Anna only thinks she has most reason to *take* the Monday option – that is, if she perceives herself to have reason for *action* – RRR doesn't apply to her either, since it only mentions perceptions of reasons for attitudes. So for RRR to explain what's wrong with Anna and Beate, it must be combined with Normative Constitutivism, with an important proviso. The proviso is that the reason-takings involved in having attitudes must be competent, since according to RRR only competent perceptions of reasons give rise to rational requirements. So I must make the case that

---

<sup>9</sup> There is no shortage of efforts to show that externalism is consistent with Enya's being doxastically justified in some way (see e.g. Goldman 1986). I cannot enter into this debate here, but if some such externalist move is successful with respect to justification, there is a good chance that it can be applied to rationality as well.

Anna and Beate *competently* take themselves to have reasons for further attitudes in virtue of their beliefs and intentions.

The crucial question is whether the disposition embedded in Intention Commitment and the Action-Intention Principle, for example, tends to hit the target – for example, that we (fallibly) take ourselves to have reason to take the means when and only when we do have reason to take the means. I think that is the case, though I cannot give a full defense here. The basic idea is that the perceptions of reasons that are embedded in our attitudes track genuine *patterns of reasons* in the favorable case. Such patterns include the transmission of a reason for an end to a means or a precondition that facilitates it, and from an action to intention to perform that action (when acting is up to the agent). For example, if you have reason to take an umbrella and your taking the umbrella hangs on your intending to do so, you have reason to intend to take the umbrella. Given the existence of such patterns, someone who perceives a reason for an end, say, must perceive a reason for the means in order to be competent with reasons.

But what if one doesn't have a reason for one's end even if one thinks so, or doesn't even think she has sufficient reason for the end? Why does Reason tell us to take the means when there is no (putative) reason to transmit to the means? The rationality of intending the means in the former kind of case is relatively easy for RRR to explain. Since recognizing dependencies among reasons is a key part of any reason-sensitive competence, a competent agent will take herself to have reason to intend the means even if she's mistaken in thinking that the facts or her evidence favor intending the end. To borrow Sosa's archery metaphor, this is akin to firing an adroit shot at the wrong target. Such a shot is to one's credit as an archer, even if one's target selection isn't. Competences for recognizing reasons for ends and recognizing patterns of reasons may come

apart, as already Aristotle noted when he distinguished between practical wisdom and mere cleverness. But that doesn't undermine the status of the latter as a reasons-sensitive competence.

The second kind of case is illustrated by my akratic intention to eat the ice cream. I argued above that even in such a case, I will think I must open the fridge door (that I have conclusive reason to do so). How can such reason-taking be competent? Clearly, these cases are tricky for RRR. They're like an archer firing a shot at what she thinks, or even knows, to be the wrong target! I think the best case for why Reason is on the side of such reason-takings may be based on the predictive significance on intention (see Scanlon 2008). Suppose the archer knows she is going to take a shot at the wrong target. She regards this as a fact about herself. Her options are doing it well, in a way that's likely to hit the target, or doing it badly, for example without paying attention to wind conditions. It seems that it is still the former that manifests her archery competence. Similarly, in the akratic intention case, I'm to some extent alienated from my intention to eat the ice cream, and regard it as a fact about myself (it is, evidently, an attitude that isn't controlled by my take on reasons). In such a case, I manifest my competence with reasons by doing what I'm going to do in a way that is responsive to reasons that would transmit to the means, were I to have sufficient reason for the end. The alternative, after all, would be doing reason-responsiveness badly, say by selecting a means that is unlikely to result in the realization of my aim. This isn't a case of recognizing transmitted reasons, but Reason telling us to make the best of a bad situation, as it were.

#### *4.3 The Authority of Rationality*

Let us finally turn to the issue of authority of rationality. I've emphasized that there are two kinds of question here, first-personal ("why should I be rational?") and third-personal (why is it a

criticism to call someone irrational?). Both questions are easily answered if rationality is deontically significant – that is, if we *ought* to do what rationality tells us to do. Indeed, for Benjamin Kiesewetter (2017), the criticism point is the main argument for strong deontic significance. He argues that we’re not criticizable for failing to F unless we ought to F, so since we *are* criticizable for failing to respond as rationality requires us to, it must be the case that we ought to F if rationality requires us to F. Similarly, from the first-person perspective, it’s a straightforward answer to the why-be-rational question if I ought to do the thing that rationality tells me to do (even if it’s not *because* rationality tells me to do it).

This poses a *prima facie* challenge for RRR, since I’ve assumed that we can have reasons we’re ignorant of (or at least misleading evidence that would lead a competent person astray), and consequently might rationally intend to F, even if it is not the case we ought to intend to F. The demands of Reason and reasons usually coincide, but not always. To be sure, if objectivism about ought is false and what we ought to do is relative to our evidence (and our reasons similarly informationally constrained), it is almost always the case that what Reason tells us to do and what we ought to do coincide – for example, in my book-borrowing case, Della ought to intend to return the book, even though it is impossible, unbeknownst to her, which coincides with what RRR says about the case. (It is only in the misleading higher-order evidence cases that the two come apart.) Under perspectivism about ought, then, rationality is standardly deontically significant when understood along the lines of RRR, and the challenge is defused.

But suppose perspectivism is false. (I will do so in the rest of this section.) Then, in the book-borrowing case as stipulated, Della ought not intend to return the book, but rationality nevertheless demands her to intend to do so, given her competent take on reasons in the face of limited access to facts. Even more worryingly, it might well be that were she to exercise her

Reason, she would end up forming the intention she ought not form, so, plausibly, she ought not exercise her Reason. Suppose, then, that she irrationally fails to form the intention to return the book. How can she be criticizable for lacking the intention, if she ought not have it, and ought not use her Reason either?

To approach this issue, let's start with a parallel with morality, where it seems clear that doing the right thing isn't enough. Imagine the following kind of rare scenario. A t-shirt manufacturer and retailer is seriously abusing workers with legal immunity, due to corruption. People ought not buy its products – indeed, stealing them happens to be the only way to force the company out of business. Now, Erika fancies one of the t-shirts. She's not aware of the abuse, and even if she were, her desire for the shirt would be stronger than her desire for justice, though she is generally a decent person. Indeed, she really wants the shirt, even though she can't at the moment afford it. By stipulation (and given objectivism about ought), she ought not pay for the shirt, and hence ought not intend to pay for it (since it's up to her whether she pays for it). And to be sure, as she hides it inside her dress, she doesn't intend to pay for it – she just intends to take it. So she has the intention she ought to have. Yet she's apparently morally criticizable. Why and for what? As I see it, the key element is that she might easily have done the wrong thing, given the way she formed her intention. Had she not got morally lucky, it would not have been the case she has the intention she morally ought to have. It's not just that her selfish motives are criticizable, but *she* is criticizable for forming the intention on their basis – she might have had those motives, and nevertheless decided otherwise. At the same time, had she manifested the virtue of honesty, she would have done the wrong thing in this very unusual circumstance, not knowing the facts, so

she ought not have been honest on this occasion.<sup>10</sup> To be sure, the fact that being honest generally results in doing the right thing is a *pro tanto* reason to be honest (and explains, in part, why honesty is a virtue), but in this case the reason is either outweighed by the considerations that favor stealing, or it is disabled in Dancy's sense.

The kind of criticism that is appropriate in cases of right attitude or action without moral worth can be called *aretaic*. As Gary Watson (1996/2004, 278) observed, while familiar forms of blame implicate emotions like resentment or indignation that are ways of holding the agent accountable for something they can control, there are also critical attitudes that are warranted toward someone just because something bad is *attributable* to her, or discloses something about who she is. Nevertheless, such attitudes involve more than mere evaluation or grading or appraisal, the kind of response that might be appropriate to a malfunctioning artefact or a non-attributable defect in a person. According to David Shoemaker's take, aretaic 'blame' consists in attitudes of disdain, which involve "feelings of superiority, thoughts about aretaic failure, and a motivation to be better than the disdained agent, or at least not to emulate him or her" (Shoemaker 2015, 42). He observes that such attitudes have a "practical pay-off": "we make all sorts of decisions about who we will interact with, or spend any significant amount of time with, or work on loving more, on the basis of such assessments" (Shoemaker 2011, 615).

If rationality is the rule of Reason (and perspectivism about ought is false), irrationality is criticizable in the same fashion without being deontically significant. According to RRR, we have a strong *pro tanto* reason to exercise and be guided by our Reason: in general, exercising and obeying our reason-sensitive competences is our *best bet* of getting it right. That's not only a reason to do what rationality tells us to do, but to do so *because* it tells us to do so. It is also an

---

<sup>10</sup> I'm assuming that if you ought to refrain from F-ing and you would F if you G-d, you ought not G (assuming you can refrain from G-ing).

answer to the general why-be-rational question from the first person perspective: “Because otherwise you’ll easily get it wrong” is a good answer to someone who asks why they should be rational, or at least aspire to be. Importantly, this kind of answer is not available to pure perceived reason or structural views, because there’s no reason to think that conforming to perceptions of reasons without regard for competence or achieving mere coherence make it likely that we get it right.<sup>11</sup> So although rationality isn’t deontically significant, it has *authority* that derives from the authority of reasons (and the value of doing what we ought to do). This is why RRR, unlike these competing non-deontic views, can explain why it is *good* to be rational, or why rationality is a kind of *virtue*.

Consider rational criticism from this perspective. Even if Della gives the response she ought to give, she does so only by dint of good luck, if she fails to exercise her Reason or abide by its demands. Given the way she formed her stance, she might easily have given the wrong response. The criticism isn’t that she *ought* to have used her Reason, but that having failed to do so shows that she adopted her stance (of not intending to return the book as promised) in an irresponsibly risky way, which is bad from the perspective of reason-responsiveness. She is criticizable because she didn’t respond to what she knows by forming the intention to return the book, even though it happens to be the case that she ought not have that intention. Again, this is not just a matter of appraising or grading Della. She hasn’t somehow malfunctioned, but has fallen short of an ideal she has reason to aspire to. She might not be to blame for this (we might blamelessly lack competence, after all), but she is liable to aretaic criticism. It is important here that such criticism need not concern someone’s long-lasting traits, but also individual acts and attitudes. Calling Della’s failure to form the intention irrational is analogous to calling someone’s

---

<sup>11</sup> The latter point is thoroughly argued by Kolodny (2008).



action cowardly, or saying that something is the honest thing to do. You don't have to be cowardly or honest to perform such acts on occasion. At the same time, on their own, cowardly actions are evidence of some defect in the agent, and honest acts show something good about the agent – though it's perfectly possible that the honest thing to do is sometimes the wrong thing to do, as in Erika's case. Similarly, saying that Della's lack of intention to return the book is irrational (or that Della is irrational in failing to intend) is saying that the absence of the intention in the presence of competent judgment to the effect she ought to have it, on its own, indicates a defect in the way she governs herself. Consequently, it invites us to adjust our opinion of her self-governance downwards, and perhaps reduce our level of trust in her, and invites her to change for the better without implying that she's blameworthy for her rational shortcomings.<sup>12</sup> In this way, irrational attitudes make us liable to criticism, even in the exceptional case in which they turn out to be the ones to have.

## 5. Conclusion

I started out with the observation that we seem to attribute irrationality on very different grounds – when someone's attitudes clash with their own view on reasons, when they somehow fail to fit together, and when they come apart from their reasons or evidence for responses. What I've attempted to do in this paper is to find a common thread that holds our thinking about rationality together. I've argued that if we start with the simple idea that the demands of rationality are those of our Reason or fallible competence to recognize favoring relations, we can get at least quite far. We can see that our subjective take on reasons will indeed be rationally authoritative when it is

---

<sup>12</sup> Such responses are the counterpart of the disdain and practical pay-off Shoemaker identified as elements of moral aretaic 'blame'. For a more detailed discussion of appropriate responses in the epistemic case, see Kauppinen (forthcoming).

competent, but won't be when it's not even minimally competent. If, as I've suggested, our commitment-like attitudes constitutively involve perceptions of reasons for further attitudes, and if there are genuine patterns of reasons that mirror these perceptions, our Reason also encourages having coherent attitudes. And finally, since Reason by definition tends to get favoring relations right, we'll tend to believe truths and do the right thing if we are ruled by our Reason, unless we're being fed systematically misleading evidence. So, on this view, there is a common source underlying the diverse manifestations of rationality and irrationality. And the fallible link to objective reasons suffices to explain why it's a good thing for us to be rational, and why falling short of the ideal merits a biting enough form of aretaic criticism. So while the argument of this paper is at times exploratory, I hope it suffices to show that the idea that rationality is the rule of reason is a promising contender.<sup>13</sup>

## References

- Bratman, Michael (1987). *Intention, Plans, and Practical Reason*. Stanford: CSLI Press.
- Broome, John (2007). Does Rationality Consist in Responding Correctly to Reasons? *Journal of Moral Philosophy* 4 (3), 349–374.
- (2013). *Rationality Through Reasoning*. Blackwell: Oxford.
- Brunero, John (2010). The Scope of Rational Requirements. *Philosophical Quarterly* 60, 28–49.

---

<sup>13</sup> This paper benefited greatly from a great number of invaluable comments and discussions. I'd like to thank especially John Broome, Krister Bykvist, Anandi Hattiangadi, Arto Laitinen, Barry Maguire, Maria Lasonen-Aarnio, Lilian O'Brien, Jonas Olson, Panu Raatikainen, Nick Schackel, Russ Shafer-Landau, Michael Smith, Jussi Suikkanen, Teemu Toppinen, Jonathan Way, and audiences in Stockholm, Helsinki, Providence, Berlin, and Cardiff. Written comments by Alex Gregory and Benjamin Kiesewetter proved particularly fruitful. Special thanks go to an anonymous referee for this journal, whose incisive comments and questions resulted in a thorough overhaul of the paper. Funding for work on this paper was provided by the Academy of Finland research project Well-Being, Agency, and Happiness (287776).

- Cohen, Stewart (1984). Justification and Truth. *Philosophical Studies* 46, 279–295.
- Dancy, Jonathan (2004). *Ethics Without Principles*. Oxford: Oxford University Press.
- Goldman, Alvin (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
- Greco, John (2010). *Achieving Knowledge*. Cambridge: Cambridge University Press.
- Kauppinen, Antti (2015). Favoring. *Philosophical Studies* 172 (7), 1953–1971.
- Kauppinen, Antti (forthcoming). Epistemic Norms and Epistemic Accountability. *Philosophers' Imprint*.
- Kavka, Gregory (1983). The Toxin Puzzle. *Analysis* 43 (1), 33–36.
- Kelly, Thomas (2014). Evidence. *Stanford Encyclopedia of Philosophy*.
- Kiesewetter, Benjamin (2017). *The Normativity of Rationality*. Oxford: Oxford University Press.
- Kim, Jaegwon (1988). What Is “Naturalized Epistemology”? *Philosophical Perspectives* 2, 381–405.
- Kolodny, Niko (2005). Why Be Rational?. *Mind* 114, 509–563.
- (2008). Why Be Disposed to Be Coherent? *Ethics* 118 (3), 437–463.
- (forthcoming). Instrumental Reasons. In Daniel Star (ed.), *The Oxford Handbook of Reasons and Normativity*. Oxford: Oxford University Press.
- Lackey, Jennifer (2008). Knowledge and Credit. *Philosophical Studies* 142 (1), 27–42.
- Lasonen-Aarnio, Maria (forthcoming). Enkrasia or Evidentialism? Learning to Love the Mismatch. *Philosophical Studies*.
- Littlejohn, Clayton (2012). *Justification and the Truth-Connection*. Cambridge: Cambridge University Press.
- Lord, Errol (forthcoming). What You’re Rationally Required to Do and What You Ought to Do (Are the Same Thing!). *Mind*.

- Parfit, Derek (2011). *On What Matters*. Oxford: Oxford University Press.
- Raz, Joseph (2005). The Myth of Instrumental Rationality. *Journal of Ethics and Social Philosophy* 1 (1), 1–28.
- Raz, Joseph (2011). *From Normativity to Responsibility*. Oxford: Oxford University Press.
- Scanlon, Thomas (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- (2007). Structural Irrationality. In Geoffrey Brennan et al (eds.). *Common Minds: Themes from the Philosophy of Philip Pettit*. Oxford: Oxford University Press, 84–103.
- Scanlon, Thomas (2008). *Moral Dimensions*. Cambridge, MA: Belknap Press.
- Schroeder, Mark (2009). Means-End Coherence, Stringency, and Subjective Reasons. *Philosophical Studies* 143 (2), 223–248.
- Shoemaker, David (2015). *Responsibility from the Margins*. New York: Oxford University Press.
- Sosa, Ernest (2007). *A Virtue Epistemology*. Oxford: Oxford University Press.
- Sosa, Ernest (2010). *Knowing Full Well*. Princeton: Princeton University Press.
- Sylvan, Kurt (2015). What Apparent Reasons Appear to Be. *Philosophical Studies* 172: 587–606.
- Wedgwood, Ralph (2002). Internalism Explained. *Philosophy and Phenomenological Research* 65 (2), 349–369.
- Worsnip, Alex (forthcoming). The Conflict of Evidence and Coherence. *Philosophy and Phenomenological Research*.